Final Project - AI in Health Technologies D

Niccolò Salvi
Aalto University, Helsinki, Finland
niccolo.salvi@aalto.fi

I. Introduction

Knee osteoarthritis (OA) is a prevalent degenerative joint disease that affects millions worldwide, leading to pain, stiffness, and reduced mobility. Accurate segmentation of knee cartilage from magnetic resonance imaging (MRI) scans is crucial for quantitative assessment of cartilage thickness and volume, which are key biomarkers for OA progression and treatment evaluation. Traditional segmentation approaches often treat the problem as a binary classification task, assigning hard labels (e.g., 0 for background, 1 for foreground) to each pixel. However, this formulation overlooks the inherent uncertainty at cartilage boundaries, where inter-observer variability and imaging artifacts can lead to ambiguous delineations.

In this work, we adapt uncertainty-aware segmentation ideas inspired by the SAUNA framework to the knee cartilage segmentation task. Specifically, we reframe the problem as a regression task using "soft" labels derived from Signed Distance Fields (SDFs). SDFs transform the binary masks into continuous values representing the signed distance to the nearest boundary, with negative values inside the object and positive outside. This encoding captures spatial uncertainty and provides richer geometric information for the model to learn.

Our approach employs an implicit neural representation to predict these continuous SDF values, trained with regression losses such as L1 loss. We compare this method against a trivial baseline using hard labels and cross-entropy loss, and we ensure patient-aware data splitting to prevent leakage.

This report outlines the methodology, experimental setup, and results, highlighting the potential of soft label regression for enhancing medical image segmentation tasks.

II. Methodology

A. Data Preprocessing and Soft-Label Transformation

We resample each 3D MRI volume into 2D axial slices and keep the original patient IDs to enforce patient-aware splits. Slice intensities are normalized per volume using percentile-based scaling (1st–99th) and then mapped to $[0, 1]$ at training time. Data augmentation is applied with paired transforms (resize, random flips, random crops) so that images and masks remain aligned.

Soft labels are generated by converting binary cartilage masks to Signed Distance Fields (SDFs). For every pixel we compute the signed Euclidean distance to the closest boundary (negative inside cartilage, positive outside). Distances are clipped to a fixed range and normalized to $[-1, 1]$ to stabilize regression. This representation provides continuous targets that encode boundary uncertainty while preserving geometric structure.

B. Baseline Model

As a baseline we use a conventional convolutional segmentation network trained on hard binary masks. The model predicts background vs. cartilage per pixel and is optimized with Cross-Entropy loss, representing the standard supervised segmentation setting used for comparison with the regression formulation.

C. INR Regression Model

The implicit neural representation (INR) is implemented as a coordinate-based MLP. It takes normalized (x, y) coordinates as input and outputs a single continuous value per coordinate, corresponding to the SDF value. The final layer is linear to support unbounded regression output, and the model is optimized to fit the continuous SDF targets.

D. Loss Functions

- **Baseline:** Cross-Entropy loss on binary labels.
- **Regression (SDF):** L1 loss between predicted SDF and ground-truth SDF.

III. Experiments

A. Experimental Setup

All experiments were implemented using PyTorch. We utilized a 5-fold cross-validation strategy to ensure the robustness of our results. Crucially, the data splitting was **patient-aware**: all slices belonging to a specific patient ID were assigned exclusively to either the training or the validation set for a given fold, preventing any data leakage.

Training details:

- **Optimizer:** AdamW with a learning rate of 1×10^{-3} and weight decay of 5×10^{-5}.
- **Batch Size:** 2 slices.
- **Epochs:** 10 epochs per fold.
- **Hardware:** Training was performed on a Google Colab GPU (T4).

B. Evaluation Metrics

To evaluate segmentation performance, we employed the Intersection over Union (IoU), also known as the Jaccard Index. For the SDF regression model, the predicted continuous map $\hat{\Phi}(x)$ was thresholded at zero to obtain the binary mask for metric calculation:

$$\text{Mask}_{\text{pred}} = \{x \mid \hat{\Phi}(x) < 0\}$$

We report the Mean IoU (average over all classes) and the Foreground IoU (specific to the cartilage class), as the background class dominates the image statistics.

C. Results and Comparison

1) Quantitative Comparison: Table I summarizes the performance averaged across the 5 folds. The standard classification baseline achieved a Mean IoU of 0.546 ± 0.007, slightly outperforming the SDF regression approach (0.537 ± 0.004).

TABLE I: Cross-validation results (Mean $\pm$ Std Dev over 5 folds). The 'Cartilage IoU' represents the foreground class performance.

Method	**Mean IoU**	**Cartilage (FG) IoU**	**Background IoU**
Baseline (Hard Mask)	$\mathbf{0.546 \pm 0.007}$	$\mathbf{0.103 \pm 0.014}$	$\mathbf{0.988 \pm 0.001}$
SDF Regression	0.537 ± 0.004	0.086 ± 0.006	0.988 ± 0.001

The Baseline model demonstrates a better capability in capturing the specific cartilage region (FG IoU 0.103 vs 0.086). This suggests that while the SDF model successfully learns the geometry (as evidenced by the high background IoU and comparable Mean IoU), the L1 loss on the distance field might be less aggressive than Cross-Entropy in penalizing errors specifically at the fine boundaries of the small cartilage structures.

IV. Conclusion

We reframed knee cartilage segmentation as a regression task by predicting signed distance fields and compared it against a standard hard-mask baseline under patient-aware cross-validation. While the regression approach provides continuous geometric targets, the baseline achieved slightly higher foreground IoU, indicating that direct classification remains more effective for fine cartilage boundaries in this setup. These findings motivate future work on loss reweighting or positional encoding for INR models to better capture high-frequency structure while preserving uncertainty-aware representations.

V. Member Contributions

I, Niccolò, am the sole contributor to this project.

Although this was intended as a group assignment, I tried to contact the other members but did not receive a response. Therefore, I carried out all the work independently.